

Annotation

By Daisy Abbott, University of Glasgow

- [Introduction](#)
- [Short-term Benefits And Long-term Value](#)
- [He/fe Perspective](#)
- [e-Science Perspective](#)
- [Issues To Be Considered](#)
- [Additional Resources](#)

1. Introduction

Annotation is the process of adding notes on or to something. It plays a particularly important role in scholarly work as explanations, descriptions, and interpretations can inform any future research on the annotated data. Annotations can be applied to all kinds of digital data and the process can be undertaken for a variety of reasons including description, correction, classification, reviewing, interpretation, and augmentation of the data.

Examples Of Annotation In Practice

Different communities of practice use annotation for different purposes and in different ways. The following list of examples is not exhaustive:

- **Annotated bibliographies** provide descriptions about how each source is useful to an author in constructing a paper or argument.
- **Linguistic annotation** can refer to any descriptive or analytic data applied to raw language data. Annotations are often textual (e.g. transcripts) but can also include any other type of data (e.g. syntactic analysis). A collection of texts with linguistic annotations is known as a **corpus**.
- Programmers often add annotations (**comments**) to source code explicating in a more easily human-readable fashion the function of the code and perhaps also ideas for development.
- **DNA annotation** has been ongoing since the 1980s in the field of molecular biology, usually in predefined fields in sequence databases.
- **Automatic image annotation** is the process by which a computer system assigns metadata in the form of keywords to a digital image. Images can also be annotated with technical information which speeds up content-based retrieval, for example histograms of colour content and segmentation information.
- Performers of all kinds often make annotations to physical objects (e.g. scores and scripts) during the rehearsal process. Digital records can also be annotated to increase understanding, for example, **video annotation** can be in the form of time stamps, overlaid text or images, and commentaries.
- Annotations can be associated with a web resource and can allow users to personalise a web page (editing, adding, or removing information) without modifying the underlying resource. Due to their connectivity, **web annotations** are particularly collaborative in nature. Many Web 2.0 applications now make extensive and imaginative use of annotations to review, rate, improve, adapt, discuss, qualify, or critique content provided by both website owners and other users.
- **Collaborative annotation** using free keyword tags to categorise content can create a user-generated taxonomy, known as a **folksonomy**. This type of **social bookmarking** is popular due to participation being fast and intuitive and can result in new ways of visualising links between information such as tag clouds.

2. Short-term Benefits And Long-term Value

- Descriptive annotations, like metadata, are crucial for information retrieval. For example, simple keyword annotations can summarise a long text, or describe non-textual media making it possible to use text searching tools to discover images, audio, or video. Many content-based retrieval systems also use non-textual annotations of one kind or another which distill various aspects of files into data which can be processed more quickly, for example music can be matched by comparing previously-extracted numerical data on volume levels and images can be rated for similarity based on histograms of colour data.[1](#)
- Annotations also augment core data by providing layers of explanation and interpretation. In this way, each contribution provides future users with data enriched by multiple opinions and several areas of expertise. Annotations can even identify errors in the core data.
- In the sciences, research is becoming increasingly reliant on the reuse of existing datasets. As part of an ongoing knowledge base, annotation is a crucial learning and interpretation tool. In some fields such as molecular biology, the major value lies in the annotations rather than the core data and this information represents a huge investment of skill and effort.[2](#)
- Annotations can assist in integrating results obtained by different research groups by providing aggregation information for multiple databases, and identifying the provenance of each part of the data.
- One of the primary purposes of annotation is to disseminate information (for example in textual criticism or cartography). However, annotations also provide supplementary information about the way in which a resource is used, the people using it, and the context of their contribution. In this way annotations support new research with a wider focus than that of the core data itself, such as research into social collaborative behaviours through the process of annotation.[3](#)
- Annotation supports collaboration, whether amongst a small team with similar expertise, or open to contributions from anyone with a connection to the Web. It can also promote increased engagement with the subject matter.
- Automatic annotation tools are a research challenge in their own right and can save a vast amount of time and effort from manual addition of useful supplementary information. This research area is important to both commercial and research institutions.[4](#)

3. He/fe Perspective

"Active reading is the combination of reading with critical thinking and learning, and is a fundamental part of education and knowledge work. Active reading involves not just reading, but also underlining, highlighting and scribbling comments, either on the text itself or in a separate notebook. Active watching is a similar concept applicable to the way we consume the material we watch. As text annotation promotes active reading, so will video annotation promote active watching."

— **Correia, N. & Chambel, T.** 'Active video watching using annotation' in *Proceedings of the seventh ACM international conference on Multimedia* (1999)

4. e-Science Perspective

"An essential requirement [within e-Research Environments] is the need to authenticate the source of the annotation and to restrict access to a particular group of trusted colleagues, for reasons of privacy, confidentiality or intellectual property. This is particularly important within e-Science, where the annotation or interpretation of the raw document or data, is often more valuable than the target of the annotation."

— **Schroeter, R.** et al. (2006). "A Synchronous Multimedia Annotation System for Secure Collaboratories" in *Proceedings of the Second IEEE International Conference on e-Science and Grid Computing*

5. Issues To Be Considered

- Although annotations sit alongside the original data, they are used to explain or interpret the data and therefore it is important to consider both the reliability of information and opinion in annotations and their authenticity and provenance. It is also important to demonstrate ways in which the annotations can be separated from the core data in order to prevent misinterpretation of the resource.
- Annotations from different sources can have different intrinsic 'value', for example, comments by a well-respected expert will be of more use to future users than a request for clarification. Should this hierarchy of value be represented and if so, how?
- Annotations are not a static resource; the body of annotations (and potentially also their relative value) changes over time. This creates issues of tracking when information is added but also provides opportunities for temporal analysis of the growth of the supplementary information.
- Different communities use the term 'annotation' differently and have widely varying associations. This lack of standardisation risks misunderstandings between collaborators. It is particularly important to define and use ontologies in some collaborative annotation activities to ensure consistency in data query and analysis.⁵
- Annotations themselves can be annotated, leading to issues of how to structure the relationships between data. Annotations can be logically grouped as branching threads or hierarchies which requires structural information. This leads to questions about how annotations and their authors can be tracked, and how the connections can be visualised by subsequent users.
- As they provide layers of additional information, information retrieval can be applied to the content of annotations as well as the underlying data.
- It is an ongoing challenge for annotation tools to handle the above issues whilst remaining user-friendly. The visualisation of annotations can be extremely important in terms of disseminating without overwhelming, demonstrating functionality, separating core data from annotations or distinguishing between 'high value' and 'low value' annotations, and allowing users to customise what information is displayed dependent on role or interest.

6. Additional Resources

- Buneman P. et al. (2005). "[Annotation in Scientific Data: a Scoping Report](#)"
 - Bird, S. & Liberman, M. (2001). "A formal framework for linguistic annotation" in *Speech Communication*, no. 33
 - Dill, S. et al. (2003). "SemTag and seeker: bootstrapping the semantic web via automated semantic annotation" in *Proceedings of the 12th international conference on the World Wide Web*, Budapest
 - Marshall, C. "[Annotation: from paper books to the digital library](#)"
 - Ruvane, M. (2005). "Annotation as Process: A Vital Information Seeking Activity in Historical Geographic Research" in *Proceedings of the 68th Annual Meeting of the American Society for Information Science and Technology (ASIST)* no. 42
 - Schreiber, A. et al. (2001). "Ontology-based photo annotation" in *Intelligent Systems (IEEE)*, vol. 16, no. 3
 - [STARS](#) (Semantic Tools for Screen Arts Research)
 - [Web Annotation Tools](#)
1. "[Information Retrieval](#)" DigiCULT Technology Watch Report 3, December 2004
 2. Buneman P. et al. "[Annotation in Scientific Data: a Scoping Report](#)"
 3. [Tags Network Narrative project](#)
 4. For example: Matellanes, A. et al. "[Creating an application for automatic annotation of images and video](#)" , *Automatic Annotation and Information Retrieval : New Perspectives* (Special Track at the 20th International Florida Artificial Intelligence Research Society Conference 2007); Wyman, S. et al. (2004). "[Automatic annotation of organellar genomes with DOGMA](#)"
 5. Cf. [The National Center for Biomedical Ontology](#)